



# 2017 생물통계학 특론

Youngjun Na, PhD



한마디로 정말 잘했네



## 우리의 연구가 답해야 하는 것

- “우리가 한 연구가 모든 돼지에게, 모든 소에게, 모든 닭에게 적용이 될 수 있는가?”라는 질문에 답해야 한다.
- The one simple elegant equation, to explain everything.



## 왜 연구자는 실험 설계와 통계를 알아야 하는가?

- 대부분의 경우 연구대상(동물) 집단의 모든 개체의(=모집단) 정보를 얻는 것은 불가능  
“경기도에 존재하는 모든 산란계의 몸무게를 알고 싶다”  
만약 모든 산란계의 몸무게를 잴다면 → 필요한 정보를 모두 알 수 있음(현실적으로 불가능)  
만약 100마리의 표본을 선택해 무게를 재면 → 한정된 정보만을 얻을 수 있음(현실적으로 가능)
- 하지만 통계를 이용한다면? → 한정된 정보를 이용해 잡지 않은 닭에 대해 말할 수 있다!
- 따라서 **실험단위(experimental unit)**라고 불리는 일부 개체들만을 선택
- 일반적으로 이 부분집합을 표본으로 이용하여 모집단의 성격을 추출(**대표표본**)



## 표본을 사용하기 전에 기억해야 할 것

1. 표본을 사용하여 모집단에 대한 절대적인 확실성을 얻는 것은 불가능하다.  
→ 통계는 가장 좋은 추측을 하기 위한 것이지만 확실성을 얻기위한 것은 아님
2. 편향성(bias)을 피하기 위해 표본은 항상 임의로 채집해야 한다.



# 가설 세우기

과학자처럼 생각하기 -- 가설, 실험, 반례

## 우리가 만나는 많은 문제들



문제



화났어?

미안해

뭐가 문제야?

답변



결과



화났어



에이~ 화났으면서



FAIL

헤어져!!



뭔가 미안한테?



진짜 미안해 ㅜㅜ



FAIL

그게 미안한 사람의 태도야?



내 의견이 말해줘야 돼?



다음에 다시 이야기 하자



FAIL

연락하자마 절대로

문제

답변

결과





## 결국 人生은 크고 작은 문제를 해결해가는 과정

: 우리는 이미 과학을 하고 있습니다

1. 궁금한 것(문제)이 생긴다 -- A라는 첨가제가 유량을 증가시킬 수 있을까?
2. 자료를 찾는다 -- A 첨가제가 유량에 미치는 영향에 대한 논문/책 (구글/도서관)
3. 만약 이미 연구가 되어 있다 → 연구의 필요성 X
4. 만약 연구가 되어 있지 않다 → 연구의 필요성 O → 가설을 세워봅시다

문제/아이디어 → **가설** → (실험 등을 통한)검증 → 논문

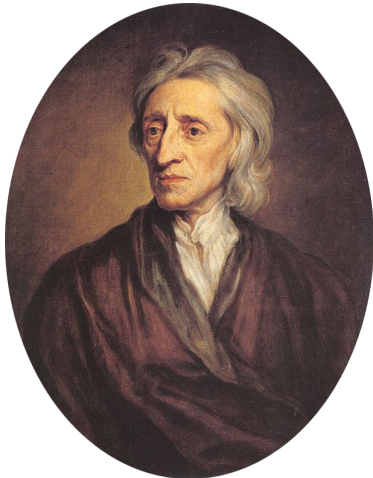


## 질문

모든 문제를 관통하는 하나의 진리가 존재하는가?

## Benedictus de Spinoza (1632-1677)

그러므로 진리는 스스로를 명백하게 드러낸다



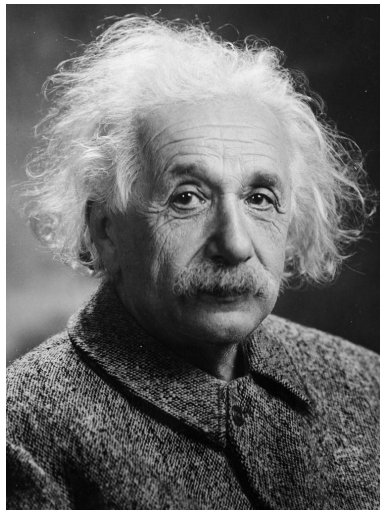
## John Locke (1632-1704)

모든 사람은 진리와 현상을 식별하는 시금석을 갖고 있다

인간은 경험을 통해 자신을 완성해 간다

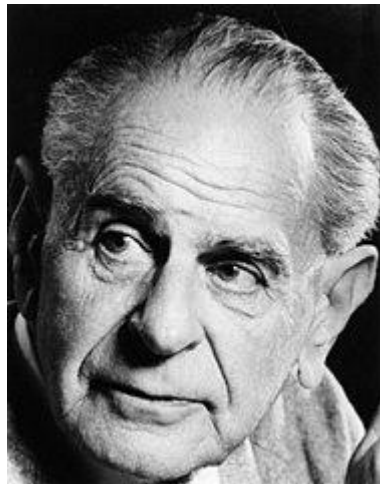
## Albert Einstein (1879-1955)

아무리 많은 실험을 하더라도 내가 옳다고 단정할 수는 없다.  
단 하나의 실험으로도 내가 틀렸다는 것이 드러날 수 있기 때문이다.



## Karl Popper (1902-1994)

- 추상적인 선을 실현하려고 하지 말고 구체적인 악을 제거하기 위해 노력하라.
- 비록 우리가 모르는 것이 너무도 많다는 사실만을 알게 된다 하더라도, 세상에 대해서 무엇인가를 배우려는 시도는 가치 있는 일이라고 생각된다. 자신의 무지를 알고 있는 이러한 상태는 우리의 여러가지 많은 어려움에 도움을 줄 수 있다. 우리가 지니고 있는 다양한 단편적인 지식들 사이에는 커다란 차이가 있으나, 우리가 끝없는 무지 속에 있다는 점에서는 모두 같다는 사실을 명심해야 할 것이다.





# 어떠한 가설/이론도 영원히 증명될 수 없다

- 미래의 어느 날 가설을 기각하거나 다른 설명으로 이끄는 증거가 발견될 수 있다

예: 천동설과 지동설

- 즉, 우리는 어떠한 가설이나 이론이 틀렸다는 것을 반증할 수 있을 뿐, **절대로 그것이 옳다는 것은 증명할 수 없다**



## 간단하게 생각해보기

- $P < 0.05$  유의차가 있다
- $0.05 \leq P < 0.10$  경향성이 있다

# 논문의 예

**Table 2.** Effect of forage to concentrate ratio on nutrient digestibility in goats (*Capra hircus hircus*) and Sika deer (*Cervus nippon hortulorum*)

| Digestibility (%) | Forage to concentrate ratio |       |       | SEM | p-value |           |
|-------------------|-----------------------------|-------|-------|-----|---------|-----------|
|                   | 25:75                       | 50:50 | 73:27 |     | Linear  | Quadratic |
| Goats             |                             |       |       |     |         |           |
| DM                | 78.1                        | 68.0  | 58.5  | 2.0 | <0.001  | 0.983     |
| OM                | 79.8                        | 69.0  | 58.6  | 2.0 | 0.002   | 0.954     |
| CP                | 80.5                        | 78.6  | 77.2  | 2.1 | 0.308   | 0.948     |
| NDF               | 47.1                        | 45.8  | 45.5  | 3.4 | 0.726   | 0.906     |
| Sika deer         |                             |       |       |     |         |           |
| DM                | 76.9                        | 63.8  | 54.5  | 0.8 | <0.001  | 0.194     |
| OM                | 78.2                        | 64.7  | 55.0  | 0.8 | <0.001  | 0.212     |
| CP                | 76.8                        | 73.2  | 72.2  | 1.3 | 0.028   | 0.368     |
| NDF               | 44.1                        | 38.2  | 40.1  | 2.5 | 0.278   | 0.253     |

SEM, standard error of the mean; DM, dry matter; OM, organic matter; CP, crude protein; NDF, neutral detergent fiber.

Na, Youngjun, Dong Hua Li, and Sang Rak Lee. "Effects of Dietary Forage-to-Concentrate Ratio on Nutrient Digestibility and Enteric Methane Production in Growing Goats (*Capra Hircus Hircus*) and Sika Deer (*Cervus Nippon Hortulorum*)."  
*Asian-Australasian Journal of Animal Sciences*, March 21, 2017.





## 검손한 가설: 귀무가설( $H_0$ , null hypothesis)

- 가설검정을 위해서는 반대되는 두 종류의 가설을 세워야 한다
  1. A라는 첨가제는 홀스타인 착유우의 유량을 증가시키는 효과가 **있다** → 대립가설(Alternative hypothesis)
  2. A라는 첨가제는 홀스타인 착유우의 유량을 증가시키는 효과가 **없다** → 귀무가설(Null hypothesis)

A 첨가제 실험결과: 효과가 있었다! vs. 효과가 없지 않았다

# 가설 검정

귀무가설: A 첨가제는 효과가 없다

| Null hypothesis (H0) is: |  |   |
|--------------------------|--|---|
| Decision                 | TRUE   | FALSE   |
| Reject H0                | <b>Type I error</b> (일반적인 p value)<br>Probability = $\alpha$ | Correct decision<br>Probability = $1 - \beta$ |
| Accept H0                | Correct decision<br>Probability = $1 - \alpha$               | <b>Type II error</b><br>Probability = $\beta$ |

즉, 귀무가설이 참(TRUE)일 경우  
가설이 기각될 확률은 5%보다 낮다( $P < 0.05$ )

## 다시 논문의 예

조농비가 증가할 경우 소화율이 감소하였다

vs.

조농비가 증가할 경우 소화율이 감소하지 않을 확률은 5% 미만이다

**Table 2.** Effect of forage to concentrate ratio on nutrient digestibility in goats (*Capra hircus hircus*) and Sika deer (*Cervus nippon hortulorum*)

| Digestibility (%) | Forage to concentrate ratio |       |       | SEM | p-value |           |
|-------------------|-----------------------------|-------|-------|-----|---------|-----------|
|                   | 25:75                       | 50:50 | 73:27 |     | Linear  | Quadratic |
| Goats             |                             |       |       |     |         |           |
| DM                | 78.1                        | 68.0  | 58.5  | 2.0 | <0.001  | 0.983     |
| OM                | 79.8                        | 69.0  | 58.6  | 2.0 | 0.002   | 0.954     |
| CP                | 80.5                        | 78.6  | 77.2  | 2.1 | 0.308   | 0.948     |
| NDF               | 47.1                        | 45.8  | 45.5  | 3.4 | 0.726   | 0.906     |
| Sika deer         |                             |       |       |     |         |           |
| DM                | 76.9                        | 63.8  | 54.5  | 0.8 | <0.001  | 0.194     |
| OM                | 78.2                        | 64.7  | 55.0  | 0.8 | <0.001  | 0.212     |
| CP                | 76.8                        | 73.2  | 72.2  | 1.3 | 0.028   | 0.368     |
| NDF               | 44.1                        | 38.2  | 40.1  | 2.5 | 0.278   | 0.253     |

SEM, standard error of the mean; DM, dry matter; OM, organic matter; CP, crude protein; NDF, neutral detergent fiber.

Na, Youngjun, Dong Hua Li, and Sang Rak Lee. "Effects of Dietary Forage-to-Concentrate Ratio on Nutrient Digestibility and Enteric Methane Production in Growing Goats (*Capra Hircus Hircus*) and Sika Deer (*Cervus Nippon Hortulorum*)."  
*Asian-Australasian Journal of Animal Sciences*, March 21, 2017.



# 어떻게 이 확률(P value)을 계산할 것인가?

## ANalysis Of VAriance

**Table 1.** ANOVA for Completely Randomized Design with  $t$  Treatments and  $r$  Replications.

| Source of variation | df       | Sum of squares | Mean squares           | F-ratio       |
|---------------------|----------|----------------|------------------------|---------------|
| Treatments          | $t-1$    | $SS_T$         | $MS_T = SS_T / t-1$    | $MS_T / MS_E$ |
| Experimental Error  | $t(r-1)$ | $SS_E$         | $MS_E = SS_E / t(r-1)$ |               |
| Total               | $rt-1$   | $SS_{Total}$   |                        |               |



# Some basic concepts

Common terms and basic concepts



# Terms

**Experimental unit:** An experimental unit is the smallest division of experimental material to which a treatment can be assigned in a single act of randomization.

**Sampling unit:** The unit of experimental material on which an observation is recorded is referred to as the sampling unit.

**Experimental error:** Results of experiments are affected not only by the action of treatments but also by unexplained, random variation, which tends to mask treatment effects. The term experimental error refers to variation that exists among experimental units treated alike.



# Terms

**Blocking:** Blocking is a technique used to group experimental units together into homogeneous groups (blocks) while allowing experimenters to use heterogeneous material. The principle objective of blocking is to reduce experimental error.

**Power:**

1. low power -- If too few experimental units are used
2. high power -- If too many experimental units are used



# Terms

**Confounding:** The word confounding means the mixing together of effects. (예: pen의 위치)

**Orthogonality:** The word orthogonality occurs in several distinct but related senses in statistics.

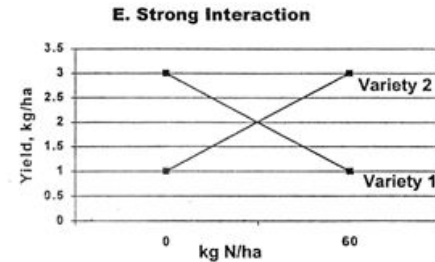
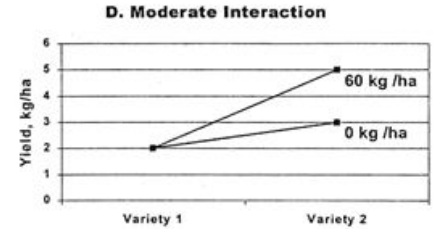
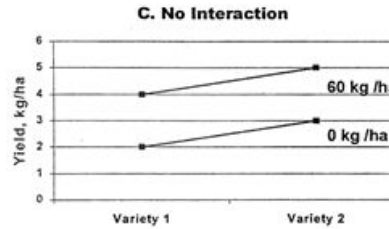
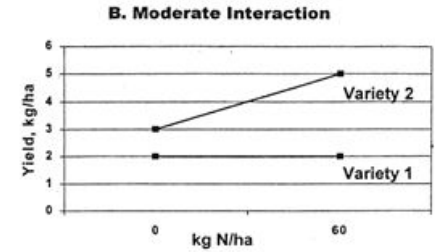
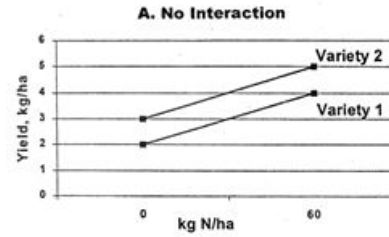
| 대조구 | 가금류 |    | 반추동물 |   |
|-----|-----|----|------|---|
|     | 산란계 | 육계 | 염소   | 양 |
| -4  | 0   | 0  | 0    | 0 |
| 0   | -1  | -1 | 1    | 1 |
| 0   | 0   | 0  | -1   | 1 |
| 0   | -1  | 1  | 0    | 0 |



# Terms

## Interaction:

Interaction means the presence of joint factor effects



**The end**